spanel: an R package to estimate the spatial panel data

Taha Zaghdoudi

Université de Jendouba, Faculté des Scieneces Juridiques Économiques et de Gestion de Jendouba

June 2016

spanel: Le package R pour l'estimation des Données de Panel Spatiale

Taha Zaghdoudi
Université de Jendouba
Faculté des Scieneces Juridiques Économiques et de Gestion de Jendouba

Juin 2015

Résumé

Ce papier introduit le package spanel développé sous le langage de programmation statistique R. spanel permet l'estimation des données de panel spatiales autoregressives on utilisant la méthode des doubles moindres carrées proposé par Kelejian and Prucha [1998], Lee [2003] et Baltagi and Liu, 2011.
Mots clés : Données de Panel ; Modèles Spatiales ; Doubles Moindres Carrés
JEL :C13,C21

spanel: An R package to estimate the Spatial Panel Data

Abstract

This paper introduces the spanel pacakge developped under the R statistical programming language.spanel estimates the spatial autoregressive panel models using the two stage least squares method proposed by Kelejian and Prucha [1998] and Lee [2003] and Baltagi and Liu, 2011.
Keywords: Panel Data; Spatial Model; Two Stage Least Squares;
JEL Classification:C13,C21

1 Méthode d'estimation

Concèderons le modèle de panel spatiale autoregressive suivant:

$$y = \lambda W y + X\beta + \mu \tag{1}$$
$$\mu = Z_\mu \mu + \nu \tag{2}$$

Où y est la variable dépendante avec une dimension $NT \times 1$,X est la matrice des variables explicatives avec une dimension $NT \times k$, β est le vecteur des paramètres de dimension $k \times 1$ et μ est le vecteur des termes d'erreur de dimension $NT \times 1$.

$W = I_T \otimes W_N$ est une matrice carrée avec W_N de dimension $N \times N$ qui représente la matrice de poids spatiale.

Avec N est le nombre des observations et T le nombre de coupes transversales. Les données sont arrangés d'abord par temps où $t = 1.....T$ ensuite par unités individuelle où $i = 1....N$.

La structure des erreurs est donnée par l'équation (2) avec $Z_\mu = \nu_T \otimes I_N$ qui est la matrice du vecteur aléatoire de l'effet individuel μ qui sont supposés indépendants et identiquement distribués (iid),$(0, \sigma_\mu^2 I_N)$. ν_T est un vecteur des perturbations de dimension $NT \times 1$ qui est supposé indépendants et identiquement distribués, $(0, \sigma_\nu^2 I_{NT})$. Aussi, μ et ν sont indépendants entre eux et par rapport à la matrice des variables explicative X.

Supposons maintenant que $A = I_T \otimes A_N$ avec $A_N = I_N - \lambda W_N$, donc l'équation (1) devient:

$$y = A^{-1}(X\beta + \mu)$$

avec:

$$E[W'_{y\mu}] = E[WA^{-1}(X\beta + \mu)\mu'] = WA^{-1}\Omega \neq 0$$

Où $\omega = E[\mu\mu']$. La variable dépendante spatialement retardée W_y est corrélée avec la perturbation μ. Donc, la méthode des moindres carrées ordinaires serait inefficace.

Supposons alors que $Z = (X, W_y)$ et $\delta = (\beta, \lambda)'$, c'est-à-dire que l'équation (1) devient:

$$y = Z\delta + \mu$$

Pour les données de panel spatiales autoregressives à coupes transversales, Kelejian and Prucha [1998] suggèrent des estimateurs de la méthode des doubles moindres carrées (2sls) basés sur des instruments faisables comme $H = (X, WX, W^2X)$ qui donnent:

$$\widehat{\delta_{2sls}} = [Z'P_H Z]^{-1} Z' P_{Hy} \tag{3}$$

Avec $P_H = H(H'H)^{-1}H'$ est la matrice de projection dans H. H peut pour le moment inclure des termes à ordre supérieur comme W^3X, W^4X,... Kelejian et al., 2004. Kelejian and Prucha [1998] montrent que les estimateurs de la méthode 2sls sont efficaces en considérant quelques condition de régularité générale.

Soit $\bar{J}_T = J_T/T$ est une matrice unitaire de dimension T. Aussi soit, $E_T = I_T - \bar{J}_T$, en suite on définie P comme la projection dans Z_μ, c'est-à-dire que $P = \bar{J}_T \otimes I_N$ et $Q = I_{NT} - P = E_T \otimes I_N$. On multiplie l'équation (3) par Q on obtient:

$$\tilde{y} = \lambda W_{\tilde{y}} + \tilde{X}\beta + \tilde{\mu} \tag{4}$$

Ceci est obtenu par le fait que $QW = (E_T \otimes I_N)(I_T \otimes W_N) = (E_T \otimes W_N) = (I_T \otimes W_N)(E_T \otimes I_N) = WQ$. On appliquant la méthode 2sls de Kelejian et Prucha (1998) a ce Q transformé des données de panel spatiales autoregressives, on obtient les estimateurs δ du 2sls effet fixe spatiale (FE-S2SLS) basé sur $\tilde{H} = (\tilde{X}, W\tilde{X}, W^2\tilde{X}) = (QX, QWX, QW^2X) = QH$.

Remarque: Le s^2 de FE-S2SLS fournit des estimateurs efficaces $\hat{\sigma_v}^2$ de σ_v^2.

Pour obtenir l'estimateur de 2sls effet between on multiplie l'équation (3) par P on aura:

$$\bar{y} = \lambda W_{\bar{y}} + \bar{X}\beta + \bar{\mu}$$

Ceci est obtenu par le fait que $PW = (\bar{J}_T \otimes I_N)(I_T \otimes W_N) = (\bar{J}_T \otimes W_N) = (I_T \otimes W_N)(\bar{J}_T \otimes I_N) = WP$. On appliquant la méthode 2sls de Kelejian et Prucha (1998) a ce P transformé des données de panel spatiales autoregressives, on obtient les estimateurs δ du 2sls effet between spatiale (BE-S2SLS) basé sur $\bar{H} = (\bar{X}, W\bar{X}, W^2\bar{X}) = (PX, PWX, PW^2X) = PH$.

Remarque: Le s^2 de BE-S2SLS fournit des estimateurs efficaces $\hat{\sigma_1}^2$ de $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$.

Baltagi (2008), montre que la matrice variance-covariance de μ est $\Omega = E(\mu'\mu) = \sigma_1^2 P + \sigma_v^2 Q$, et $\Omega^{-1/2} = (\sigma_1^{-1} P + \sigma_v^{-1} Q)$. Si on multiplie à gauche l'équation (3) par $\Omega^{-1/2}$ on aura:

$$\dot{y} = \dot{Z}\delta + \dot{\mu} \tag{5}$$

Où $\dot{y} = \Omega^{-1/2}y$, $\dot{\mu} = \Omega^{-1/2}\mu$ et $\dot{Z} = \Omega^{-1/2}Z = \Omega^{-1/2}(X, W_N) = (\dot{X}, \dot{W})$. Ceci est obtenu par $\Omega^{-1/2}W_N = (\sigma_1^{-1}PW + \sigma_v^{-1}QW) = (\sigma_1^{-1}WP + \sigma_v^{-1}WQ) = W_N\Omega^{-1/2}$. Par la suite l'équation (5) devient:

$$\dot{y} = \lambda W_{\dot{y}} + \dot{X}\beta + \dot{\mu}$$

On appliquant la méthode 2sls de Kelejian and Prucha [1998] à la transformation $\Omega^{-1/2}$ des données de panel spatiales autoregressive on obtient les estimateurs δ de l'effet aléatoire des doubles moindres carrées (RE-2SLS) données par:

$$\widehat{\delta_{RE-2sls}} = [\dot{Z}'P_H\dot{Z}]^{-1}\dot{Z}'P_{H\dot{y}}$$

avec $\dot{H} = (\dot{X}, W\dot{X}, W^2\dot{X}) = (\Omega^{-1/2}X, W\Omega^{-1/2}X, W^2\Omega^{-1/2}X) = (\Omega^{-1/2}X, \Omega^{-1/2}WX, \Omega^{-1/2}W^2X) = \Omega^{-1/2}H$

2 Exemple

Dans cette section on donne un exemple pratique de l'utilisation du package spanel.

```
# Ouvrir le pacakge
library(spanel)
# Importer les données
data(Produc)
```

```
        #Extraire la matrice de poids spatiales
data("usaww")
```

```
# Estimation de l'effet fixe
fx<-span(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,Produc,
         usaww,n=48,t=17,model="fe")
summary(fx)
```

```
## Formula:log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp
##
## Balanced Panel: n: 48 t: 17 N: 816
## Rsquared : 0.02005736
##              Estimates     Std.Err T-value   P-Value
## log(pcap) -0.04007343  0.02667848 -1.5021    0.1335
## log(pc)    0.22074355  0.02506409  8.8072 < 2.2e-16 ***
## log(emp)   0.67066333  0.03072264 21.8296 < 2.2e-16 ***
## unemp     -0.00474157  0.00091041 -5.2082 2.420e-07 ***
##            0.18718963  0.02587428  7.2346 1.083e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Le modèle between
be<-span(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,Produc,
         usaww,n=48,t=17,model="be")
summary(be)
```

```
## Formula:log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp
##
## Balanced Panel: n: 48 t: 17 N: 816
## Rsquared : 0.9939232
##               Estimates    Std.Err T-value   P-Value
## (Intercept)  1.6928901  0.7781358  2.1756 0.0298754 *
## log(pcap)    0.1723945  0.1616659  1.0664 0.2865771
## log(pc)      0.3016717  0.0910694  3.3125 0.0009655 ***
## log(emp)     0.5843175  0.1313312  4.4492 9.823e-06 ***
## unemp       -0.0026183  0.0227603 -0.1150 0.9084439
##             -0.0093646  0.0534176 -0.1753 0.8608810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Estimation l'effet aléatoire
ran<-span(log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp,Produc,
          usaww,n=48,t=17,model="re")
summary(ran)
```

```
## Formula:log(gsp) ~ log(pcap) + log(pc) + log(emp) + unemp
##
## Balanced Panel: n: 48 t: 17 N: 816
## Rsquared : 0.950528
##                Estimates     Std.Err T-value   P-Value
## (Intercept)  1.92719851  0.17102799 11.2683 < 2.2e-16 ***
## log(pcap)   -0.00392653  0.02564756 -0.1531    0.8784
## log(pc)      0.25139222  0.02381932 10.5541 < 2.2e-16 ***
## log(emp)     0.70471904  0.02898067 24.3169 < 2.2e-16 ***
```

```
## unemp       -0.00558438  0.00090403 -6.1772 1.032e-09 ***
##              0.10301823  0.02025889  5.0851 4.566e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remarque: la dernière ligne des résulats d'estimations représente W*rho.

```
# Le test de Hausman
hausman(fx, ran)

##
## Hausman Test
## Chisq: 28.51944
## P-value: 2.880384e-05
## df: 5
```

3 conclusion

La méthode des doubles moindres carrées utilisée dans ce package est la plus facile à appliquer. En effet, d'autre méthodes sont aussi utilisées pour estimer les données de panel spatiales principalement celle des maximums de vrais-semblances et des moments généralisés en particulier dans le package splm sur R et xsmle sur STATA.

References

Badi H Baltagi and Long Liu. Instrumental variable estimation of a spatial autoregressive panel model with random effects. *Economics Letters*, 111(2):135–137, 2011.

Harry H Kelejian and Ingmar R Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

Harry H Kelejian, Ingmar R Prucha, and Yevgeny Yuzefovich. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. *Advances in Econometrics: Spatial and Spatio-Temporal econometrics*, pages 163–198, 2004.

Lung-fei Lee. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4):307–335, 2003.